



University of
Sheffield

Short course created by Robert Smith, University of Sheffield & Dark Peak Analytics, for: OPENFEST – 2023



GitHub for academic collaboration

Dr. Robert Smith

Wellcome Funded Post-Doc, ScHARR, robert.smith@sheffield.ac.uk
Director, Dark Peak Analytics, rsmith@darkpeakanalytics.com





Who are we anyway?



Dr. Paul Schneider



Dr. Robert Smith



Dr. Sarah Bates



ShangShang Gu



Wael Mohammed

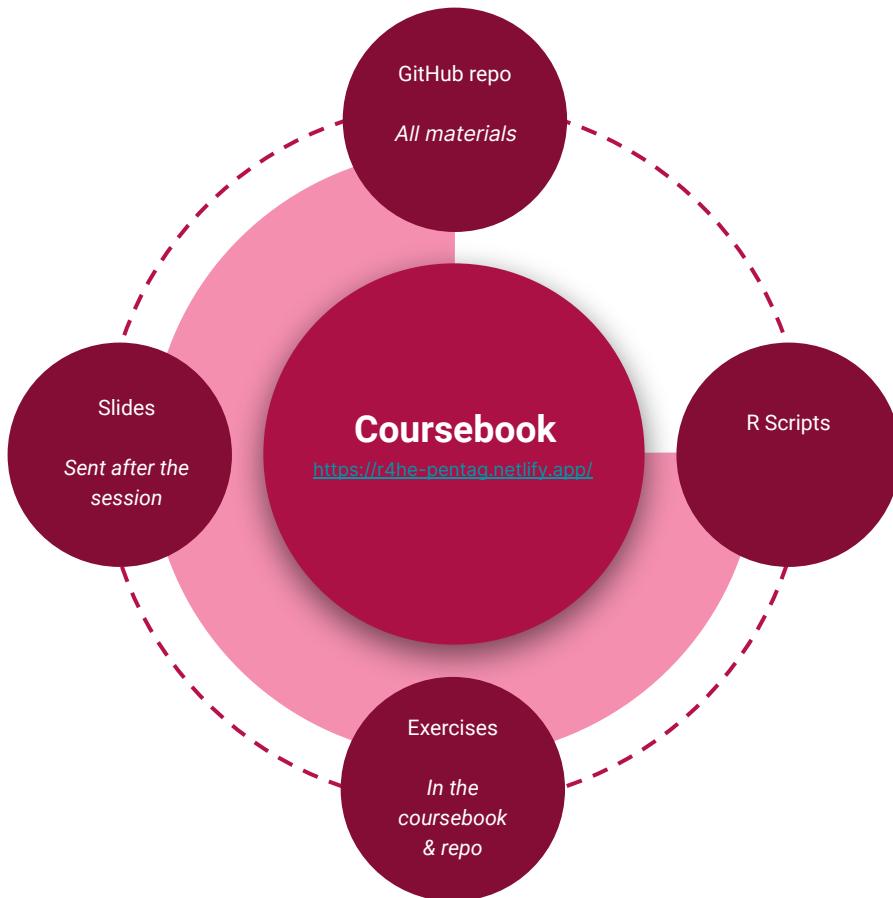




Table of contents

- [1 Using this book](#)
- [2 About the authors](#)
- [3 Background](#)
- [4 Introduction to R](#)
- [5 Version Control](#)
- [6 Intermediate R](#)
- [7 Partitioned Survival Models](#)
- [8 State Transition Models](#)
- [9 Automated Reporting](#)
- [10 R shiny for Health Economics](#)
- [11 Advanced Data Visualisation](#)
- [12 Further Resources](#)
- [References](#)

1 Using this book

Note from the authors: This work is a living document and is being adapted all the time based upon comments. New sections are being included prior to teaching. If you have suggestions for improving this book, please contact Robert Smith by email:
rsmith@darkpeakanalytics.com.

— Dark Peak Analytics Teaching Team

This book was created to provide additional support for the taught course. It includes all of the code chunks, exercises and solutions which we cover in the taught sessions. It serves as a first point of reference, and directs the reader to additional resources. The book is written in R using the `bookdown` package, which converts each chapter from an Rmarkdown file into a PDF or HTML book. Later in the course, we will cover how this process works.

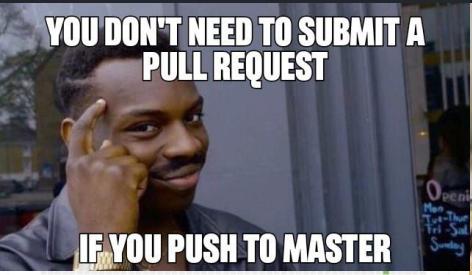
There are currently nine chapters to the book, although the content that is visible to you will depend on the scope of the course. The full list of book chapters can be found below.

Number	Topic
1.1	Introduction to R
1.2	Version Control
1.3	R project workflow
1.4	Intermediate R
1.5	Partitioned Survival Models
1.6	State Transition Models
1.7	Automated Reporting
1.8	R shiny for Health Economics
1.9	Advanced data visualization in R

We hope you find the content useful, to get in contact with any questions or queries please email Robert at rsmith@darkpeakanalytics.com. To find out more about Dark Peak Analytics see our [website](#).



Version control is hard!



Vs The guy she told you
not to be worried about



GIT PUSH

AND RUN



I HATED GIT





Version Control in RStudio Timetable

Time	Topic
10:00	Introductions
10:15	Session 1 - Presentation The concept behind Git & GitHub, why bother? Q&A on concepts and benefits of open science with R, Git & GitHub
10:45	Break
11:00	Session 2 - Demonstration & Exercises Working together using RStudio, Git and GitHub. Some hands on work, some trial and error. Q&A on practical issues with RStudio, Git & GitHub
11:50	Question and Answer
12:00	Close



Session 1 - Presentation

*The concept behind Git &
GitHub, why bother?*

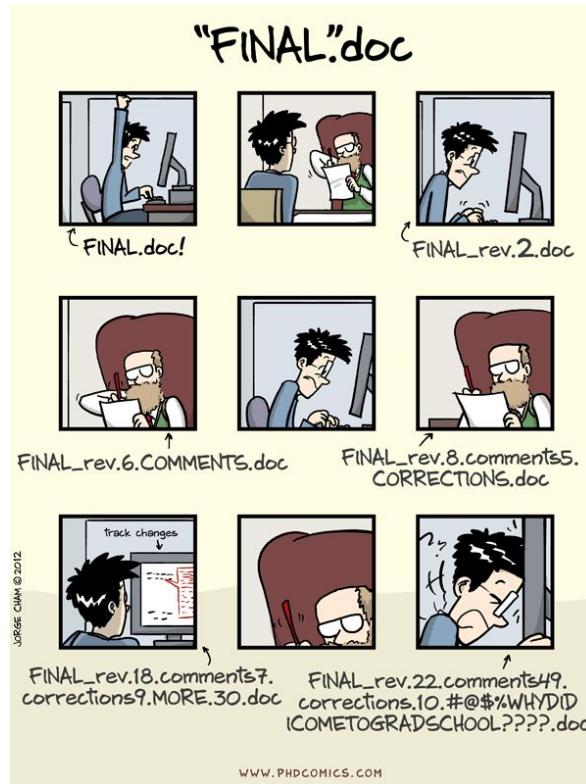


The concept behind Git & GitHub, why bother?

- What is Git?
- What is GitHub?
- GitHub for open source publishing
- GitHub for project management
- Application to Health Economics & Decision Science
 - Project model code
 - Packages & software development
 - Transparent model development, review and technical engagement
- Personal Experiences

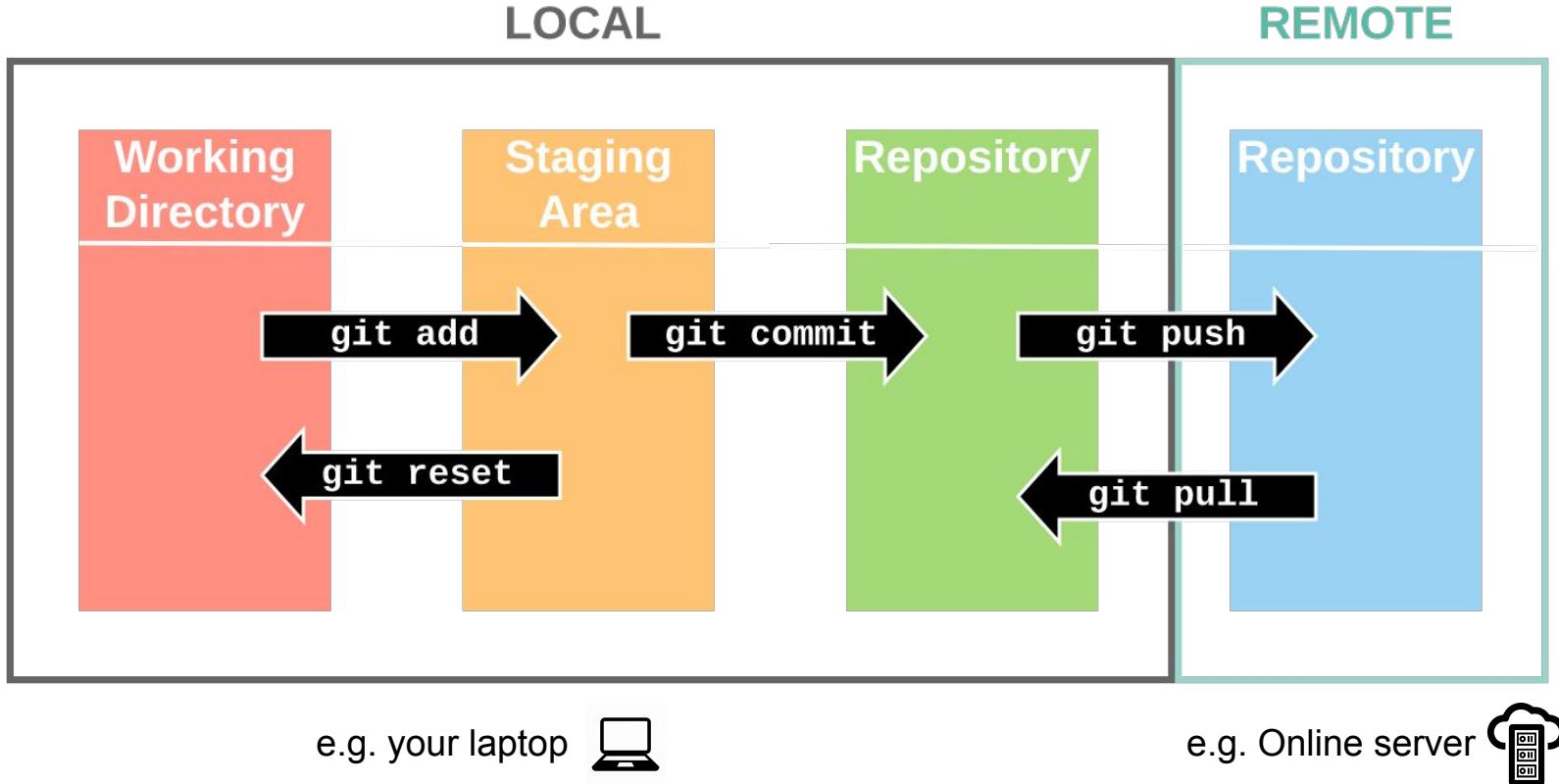


What is Git?





What is Git?





What is GitHub?

“GitHub, Inc., is an Internet hosting service for software development and version control using Git. It provides the distributed version control of Git plus access control, bug tracking, software feature requests, task management, continuous integration, and wikis for every project.^[4] Headquartered in California, it has been a subsidiary of Microsoft since 2018.^[5]

It is commonly used to host open source software development projects.^[6] As of June 2022, GitHub reported having over 83 million developers^[7] and more than 200 million repositories,^[8] including at least 28 million public repositories.^[9]

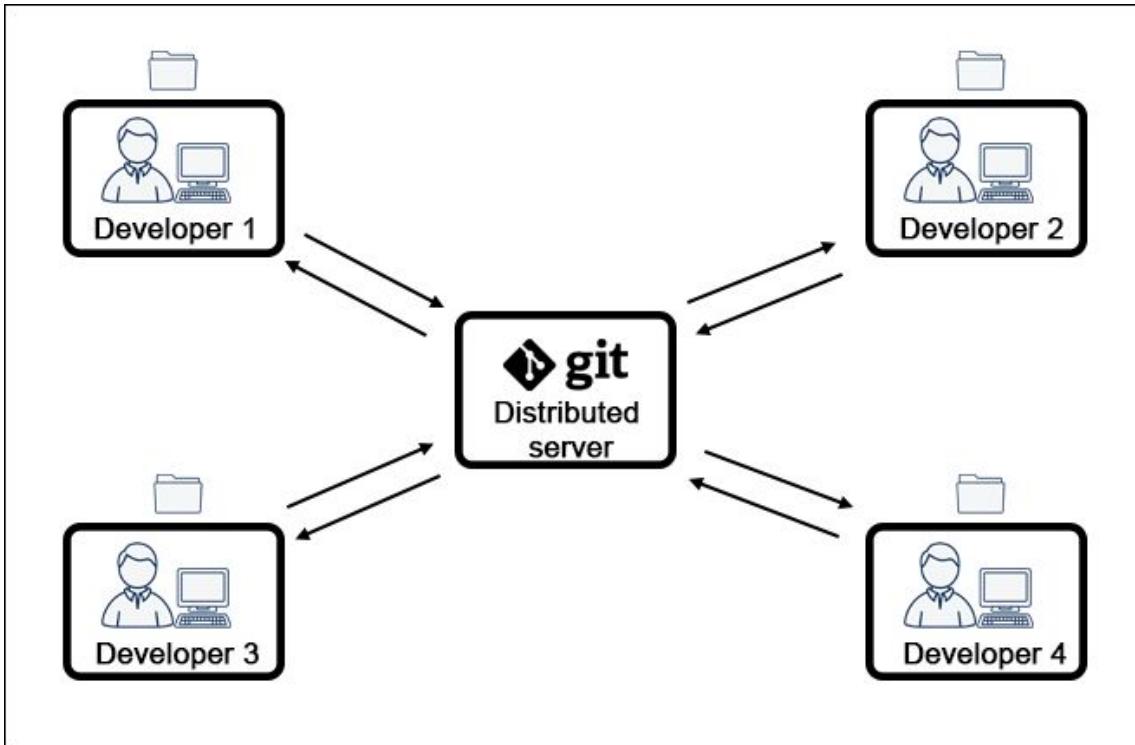
It is the largest source code host as of November 2021.”

<https://en.wikipedia.org/wiki/GitHub>





What is GitHub?



[GitHub], huh, yeah.

What is it good for?

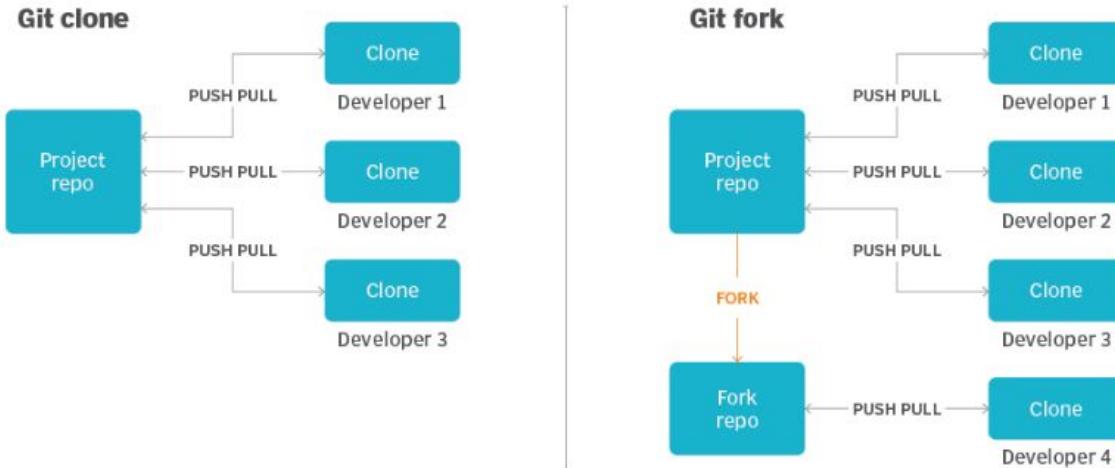
Absolutely [loads of stuff].

Say it again, y'all.



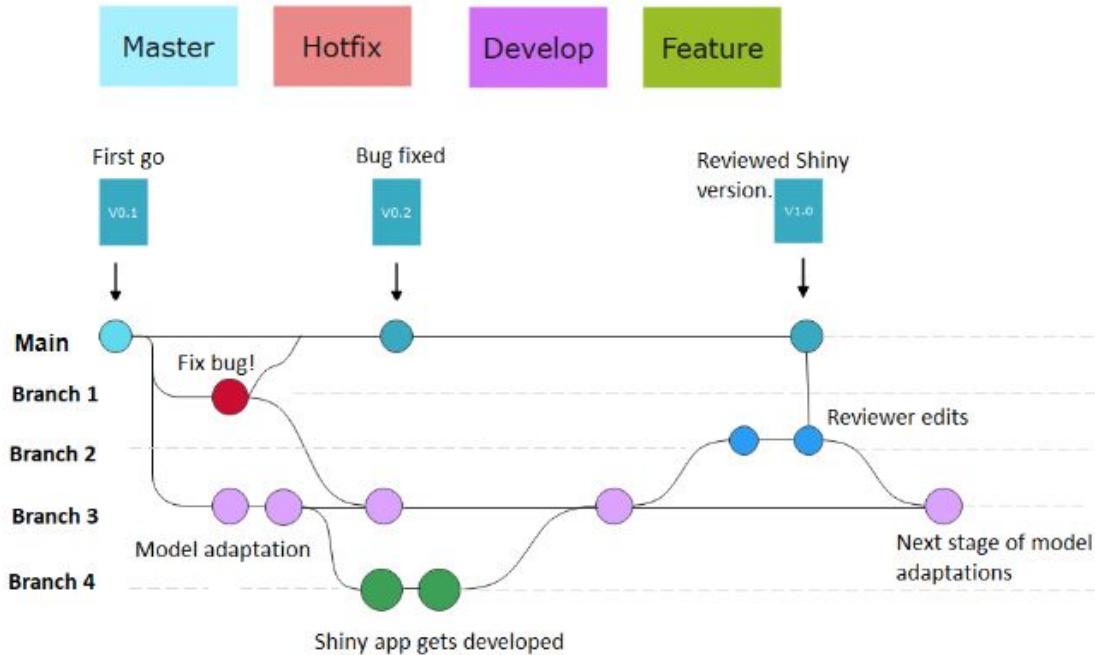
What is GitHub?

Teams who want to work closely together will tend to clone, those who may want to work independently (e.g. academics) will tend to fork. We recommend that new starters fork, while experienced teams working together clone.





Git & GitHub for project management



The diagram below shows V0.1, 0.2 and 1.0 of a health economic model.

- One *branch* fixes a bug. It gets merged back in.
- One *branch* continues with adaptations, the bug fix is incorporated.
- One *branch* adds a shiny app, it gets merged back into the adaptation
- A *reviewer* edits the model and merges back to main.



GitHub for open source publishing

Wellcome Open Research Wellcome Open Research 2020, 5:9 Last updated: 05 July 2020

RESEARCH ARTICLE



Does ethnic density influence community participation in mass participation physical activity events? The case of parkrun in England [version 2; peer review: 3 approved]

Robert A. Smith¹, Paul P. Schneider¹, Alice Bullock^{1,2}, Steve Haake^{1,2}, Helen Quirk³, Rami Cosulich¹, Elizabeth Goyder¹

¹School of Health and Related Research, University of Sheffield, Regent Court, Sheffield, S1 4DA, UK
²Advanced Wellbeing Research Centre, Sheffield Hallam University, Olympia Legacy Park, Sheffield, S9 3TU, UK

 First published: 18 Jun 2020, 5:9
<https://doi.org/10.12688/wellcomeopenres.15657.1>
Latest published: 18 Jun 2020, 5:9
<https://doi.org/10.12688/wellcomeopenres.15657.2>

Abstract
parkrun has been successful in encouraging people in England to participate in their weekly 5km running and walking events. However, there is substantial heterogeneity in parkrun participation rates across the country. After controlling for travel distances, deprived communities have significantly lower participation rates.

Results
We found significant associations between previous findings by investigating disparities in parkrun participation by ethnic density. We combined geo-spatial data available through the Office for National Statistics and parkrun data to build individual-level Poisson regression models to study the effect of ethnic density on participation rates at the Lower layer Super Output Level.

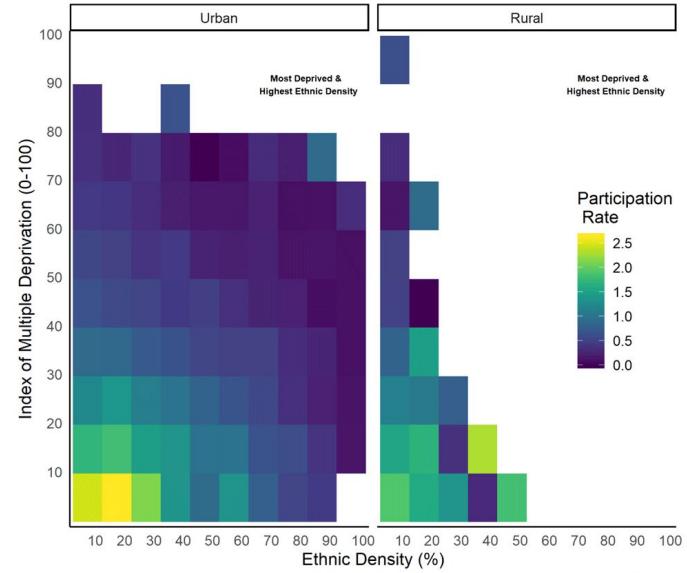
Conclusion
Community members from higher ethnic density have lower participation rates. This effect is independent of deprivation.

Conclusions: An opportunity exists for parkrun to engage with these communities and reduce potential barriers to participation.

Keywords
parkrun, Physical Activity, Ethnic Density, Deprivation

1. Anne Grunstein  The Australian Prevention Partnership Centre, NSW, Australia
University of Sydney, Sydney, Australia
2. Geddie Osakeni  University of St Andrews, St Andrews, UK
3. Stephen Senn  Consultant Statistician, Edinburgh, UK
Any reports and responses or comments on the article can be found at the end of the article.

Page 1 of 19



Smith, R., Schneider, P., et al. 2020. Does ethnic density influence community participation in mass participation physical activity events? The case of parkrun in England. *Wellcome Open Research*, 5. <https://doi.org/10.12688%2Fwellcomeopenres.15657.2>

https://github.com/ScHARR-PHEDS/DoPE_Public





GitHub for open source publishing

Wellcome Open Research | Wellcome Open Research 2020, 5:9 Last updated: 05 July 2022

RESEARCH ARTICLE

 Does ethnic density influence community participation in mass participation physical activity events? The case of parkrun in England [version 2; peer review: 3 approved]

Robert A. Smith¹, Paul P. Schneider², Alice Bullas³, Steve Haake⁴, Helen Quirk⁵, Rami Cosulich¹, Elizabeth Goyder¹

¹School of Health and Related Research, University of Sheffield, Regent Court, Sheffield, S1 4DA, UK
²Advanced Wellbeing Research Centre, Sheffield Hallam University, Olympia Legacy Park, Sheffield, S9 3TU, UK

 First published: 18 Jun 2020, 5:9
<https://doi.org/10.12688/wellcomeopenres.15657.1>
Latest published: 18 Jun 2020, 5:9
<https://doi.org/10.12688/wellcomeopenres.15657.2>

Abstract
Background: parkrun has been successful in encouraging people in England to participate in their weekly 5km running and walking events. However, there is substantial heterogeneity in parkrun participation across different communities in England: after controlling for travel distances, deprived communities have significantly lower participation rates.

Results: This study expands on previous findings by investigating disparities in parkrun participation by ethnic density. We combined geo-spatial data available through the Office for National Statistics with participation data provided by parkrun, and fitted multivariable Poisson regression models to study the effect of ethnic density on participation rates at the Lower layer Super Output Level.

Conclusions: An opportunity exists for parkrun to engage with these communities and reduce potential barriers to participation.

Keywords
parkrun, Physical Activity, Ethnic Density, Deprivation

Open Peer Review

Approval Status	1	2	3
version 2 (revision)	✓	✓	✓
18 Jun 2022	↑	↑	↑
version 1	✓	✓	?
18 Jun 2020	↑	↑	↑

1. Anne Grunstein¹ The Australian Prevention Partnership Centre, NSW, Australia
2. Gedde Osakwe² University of St Andrews, St Andrews, UK
3. Stephen Senn³ Consultant Statistician, Edinburgh, UK
Any reports and responses or comments on the article can be found at the end of the article.

Page 1 of 19

Smith, R., Schneider, P., et al. 2020. Does ethnic density influence community participation in mass participation physical activity events? The case of parkrun in England. *Wellcome Open Research*, 5. <https://doi.org/10.12688%2Fwellcomeopenres.15657.2>

Submitted by rasmith3

Aug. 4, 2022, 9:18 a.m.

Does ethnic density influence community participation in mass participation physical activity events?

Robert A. Smith, Paul P. Schneider, Alice Bullas, Steve Haake, Helen Quirk, Rami Cosulich, Elizabeth Goyder

Smith RA, Schneider PP, Bullas A et al. Does ethnic density influence community participation in mass participation physical activity events? The case of parkrun in England [version 2; peer review: 3 approved]. *Wellcome Open Res* 2020, 5:9 (<https://doi.org/10.12688/wellcomeopenres.15657.2>)

DOI: [10.12688/wellcomeopenres.15657.2](https://doi.org/10.12688/wellcomeopenres.15657.2)   

Mean reproducibility score: 10.0/10 | Number of reviews: 2

Brief Description

Background: parkrun has been successful in encouraging people in England to participate in their weekly 5km running and walking events. However, there is substantial heterogeneity in parkrun participation across different communities in England: after controlling for travel distances, deprived communities have significantly lower participation rates.

Methods: This paper expands on previous findings by investigating disparities in parkrun participation by ethnic density. We combined geo-spatial data available through the Office for National Statistics with participation data provided by parkrun, and fitted multivariable Poisson regression models to study the effect of ethnic density on participation rates at the Lower layer Super Output Level.

Results: We find that areas with higher ethnic density have lower participation rates. This effect is

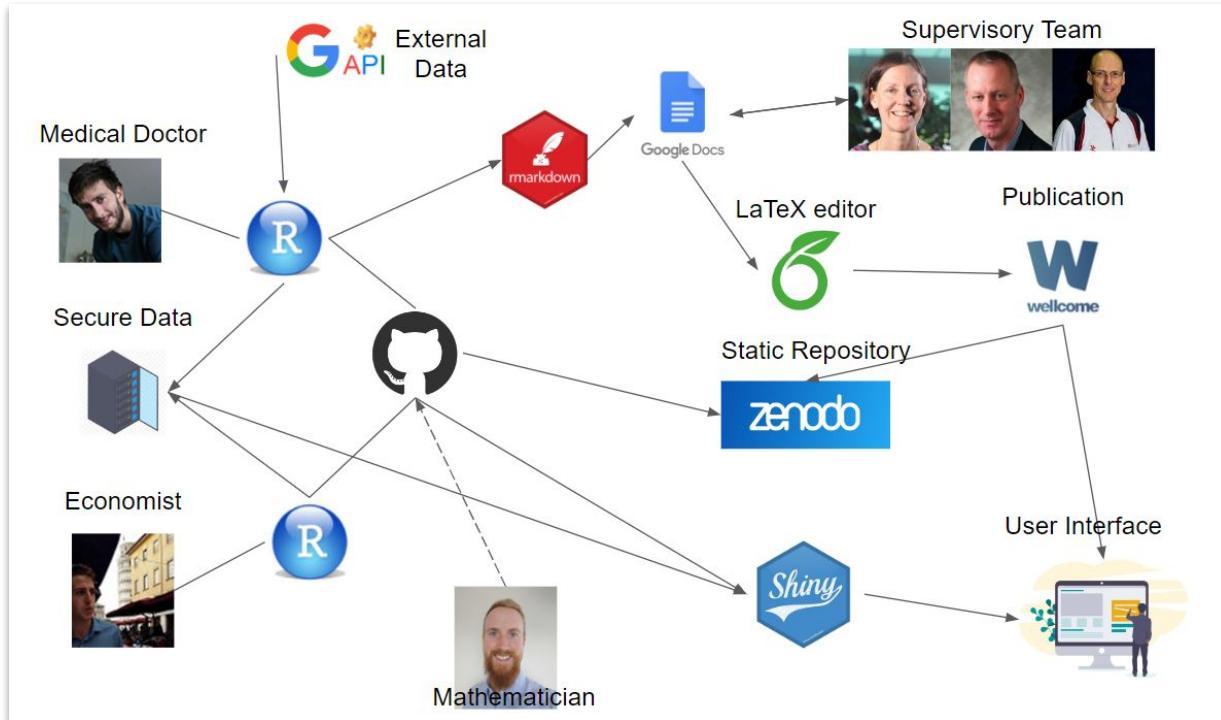
<https://www.reprohack.org/paper/79/>

https://github.com/ScHARR-PHEDS/DoPE_Public





Example case study: Academic Collaboration





GitHub for open source publishing

Wellcome Open Research Wellcome Open Research 2020, 5:69 Last updated 05 July 2022

METHOD ARTICLE

Making health economic models Shiny: A tutorial

[version 2; peer review: 2 approved]

Robert A. Smith , Paul Schneider

School of Health and Related Research, University of Sheffield, Regents Court, Sheffield, S1 4DA, UK

Equal contributors

First published: 14 Apr 2020, 5:69
Latest published: 31 Jul 2020, 5:69
<https://doi.org/10.12688/wellcomeopenres.15807.2>

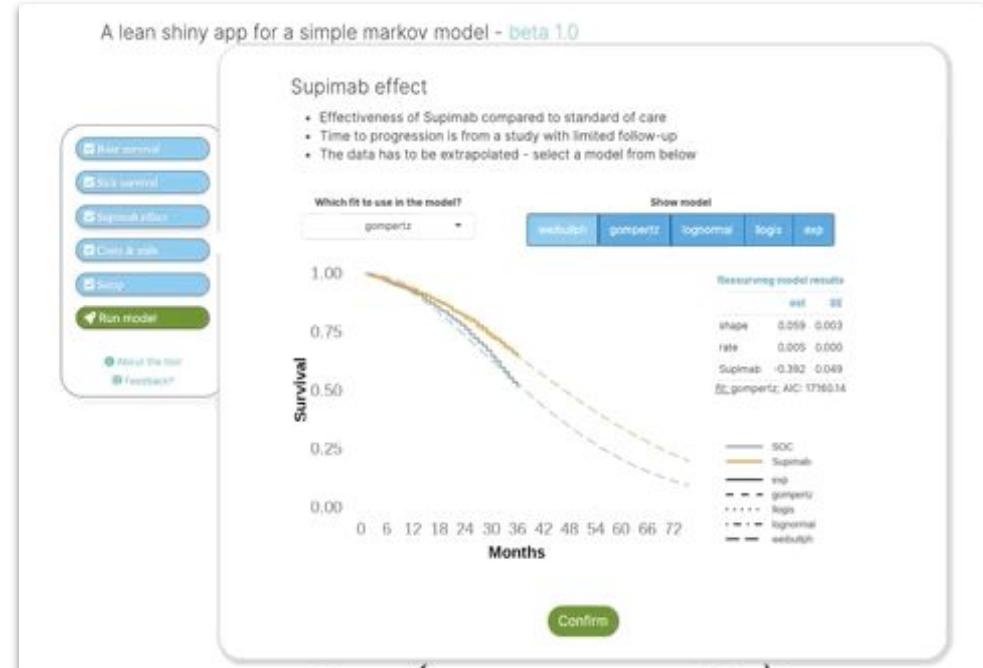
Abstract
 Health economic evaluation models have traditionally been built in Microsoft Excel, but more sophisticated tools are increasingly being used as model complexity and computational requirements increase. Of all the programming languages, R is most popular amongst health economists due to its ease of use, availability of packages and is highly flexible. However, even with an integrated development environment such as R Studio, R lacks a simple point and click user interface and lacks the ability to programme directly. This might mean the switch from Microsoft Excel to R seems daunting, and it might make it difficult to directly communicate results with decision makers and other stakeholders.

The R package Shiny has the potential to resolve this limitation. It allows programmes to be run in a web browser and developed in R with an interactive web-based user interface. Users can specify their own assumptions about model parameters and run different scenario analyses, which, in the case of regular a Markov model, can be computed within seconds. This paper provides a detailed guide to developing a Shiny application for a simple Markov model. We use a four-state Markov model developed by the Decision Analysis in Health (DARTH) group as a case study to demonstrate main principles and basic functionality.

A more extensive tutorial, all code, and data are provided in a [GitHub repository](#).

Keywords
 Health Economics, R, RShiny, Decision Science

Page 1 of 26



Smith RA and Schneider PP. Making health economic models Shiny: A tutorial. Wellcome Open Res 2020, 5:69.
<https://doi.org/10.12688/wellcomeopenres.15807.2>

<https://darkpeakanalytics.shinyapps.io/sadm-mk2/>

https://github.com/RobertASmith/paper_makeHEshiny



GitHub for open source publishing

Wellcome Open Research

Wellcome Open Research 2022, 7:194 Last updated: 24 OCT 2022

METHOD ARTICLE

Living HTA: Automating Health Economic Evaluation

with R [version 2; peer review: 2 approved]

Robert A. Smith^{1,2}, Paul P. Schneider^{1,2}, Wael Mohammed^{1,3}

¹School of Health and Related Research, University of Sheffield, Sheffield, S1 4DA, UK
²Luminary, Sheffield, S1 2GQ, UK
³Dark Peak Analytics, Sheffield, S11 7BA, UK

V2 First published: 21 Jul 2022, 7:194
Latest published: 11 Oct 2022, 7:194
<https://doi.org/10.12688/wellcomeopenres.17933.2>

Abstract
Background: Requiring access to sensitive data can be a significant obstacle for the development of health models in the health Economics & Outcomes Research (HEOR) setting. We demonstrate how health economic evaluation can be conducted with minimal transfer of data between parties, while automating reporting as new information becomes available.

Methods: We developed an automated analysis and reporting pipeline for health economic modelling, making the source code openly available on GitHub. The pipeline consists of three parts: An economic model is constructed by the consultant using pseudo data. On the data-owner side, an application programming interface (API) is hosted on a server. This API hosts all sensitive data, so no sensitive data needs to be provided to the consultant. An automated workflow is created, which calls the API, retrieves results, and generates a report.

Results: The use of modern data extraction tools and practices allows analysis of data without the need for direct access – negating the need to send sensitive data. In addition, the entire workflow can be largely automated: the analysis can be scheduled to run at defined time points (e.g. monthly), or triggered by an event (e.g. an update to the underlying data or model code); results can be generated automatically and then be exported into a report. Documents no longer need to be revised manually.

Conclusions: This example demonstrates that it is possible, within a HEOR setting, to separate the health economic model from the data, and automate the main steps of the analysis pipeline.

Keywords:
 HEOR, HTA, APIs, R, plumber

Page 1 of 27

Introduction

Living HTA - Demo Shiny App

This example application was created for the R-HTA workshop, held at Oxford University on Thursday 18th May 2023.

This app showcases the methods described in Smith, Schneider & Mohammed (2022).

The API Key is the doi of the paper: <https://doi.org/10.12688/wellcomeopenres.17933.2>

Living HTA: Automating health economic evaluation with GitHub Actions & plumber APIs

Robert Smith^{1,2}, Paul Schneider^{1,2} & Wael Mohammed^{1,3}

¹University of Sheffield, Sheffield, UK
²Luminary, Sheffield, UK
³Dark Peak Analytics, Sheffield, UK

Background
 The process of updating economic models is time-consuming and expensive, and often involves the transfer of sensitive data between parties. Here, we demonstrate how HEOR can be conducted in a way that allows clients to retain full control of their data, while automating reporting as new information becomes available.

Methods
 We developed an automated analysis and reporting pipeline for health economic modelling and made the source code openly available on a GitHub repository. It consists of three parts:

- An economic model is constructed by the consultant using pseudo data (i.e. random data, which has the same format as the real data).
- On the company side, an application programming interface (API), generated using the R package plumber, is hosted on a server. An automated workflow is created. This workflow sends the economic model to the company API. The model is then run within the company server. The results are sent back to the consultant, and a (PDF) report is automatically generated using RMarkdown.
- This API hosts all sensitive data, so that data does not have to be provided to the consultant.

Results & Discussion
 The method is relatively complex, and requires a strong understanding of R, APIs, RMarkdown and GitHub Actions. However, the end result is a process, which allows the consultant to conduct health economic (or any other) analyses on company data, without having direct access – the company does not need to share their sensitive data. The workflow can be scheduled to run at defined time points (e.g. monthly), or when triggered by an event (e.g. an update to the underlying data or model code). Results are generated automatically and wrapped into a full report. Documents no longer need to be revised manually.

Why does this matter?

By Robert J. Smith, Paul P. Schneider & Wael Mohammed

Smith RA, Schneider PP and Mohammed W. Living HTA: Automating Health Economic Evaluation with R [version 2; peer review: 2 approved]. Wellcome Open Res 2022, 7:194
<https://doi.org/10.12688/wellcomeopenres.17933.2>

https://darkpeakanalytics.shinyapps.io/living_HTA_demo/

<https://github.com/RobertASmith/plumberHE>



Application to HEDS: Project model code

Value in Health
Volume 25, Issue 3, March 2022, Pages 409-418

Economic Evaluation
CDX2 Biomarker Testing and Adjuvant Therapy for Stage II Colon Cancer: An Exploratory Cost-Effectiveness Analysis

Fernando Alarid-Escudero PhD¹, Deborah Schrag MD, MPH², Karen M. Kuntz ScD³

Show more ▾
+ Add to Mendeley Share Cite

<https://doi.org/10.1016/j.jval.2021.07.019> Get rights and content

Abstract

Objectives

Adjuvant chemotherapy is not recommended for patients with average-risk stage II (T3N0) colon cancer. Nevertheless, a subgroup of these patients who are CDX2-negative might benefit from adjuvant chemotherapy. We evaluated the cost-effectiveness of testing for the absence of CDX2 expression followed by adjuvant chemotherapy (fluorouracil combined with oxaliplatin [FOLFOX]) for patients with stage II colon cancer.

Methods

We developed a decision model to simulate a hypothetical cohort of 65-year-old patients with average-risk stage II colon cancer with 7.2% of these patients being CDX2-negative under 2 different interventions: (1) test for the absence of CDX2 expression followed by adjuvant chemotherapy for CDX2-negative patients and (2) no CDX2 testing and no adjuvant chemotherapy for any patient. We derived disease progression parameters, adjuvant chemotherapy effectiveness and utilities from published analyses, and cancer care costs from the Surveillance, Epidemiology, and End Results (SEER)-Medicare data. Sensitivity analyses were conducted.

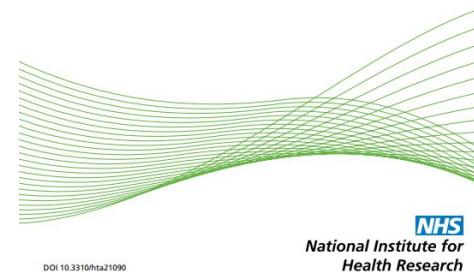
HEALTH TECHNOLOGY ASSESSMENT

VOLUME 21 ISSUE 9 MARCH 2017
ISSN 1366-5278

Check for updates

Oral anticoagulants for primary prevention, treatment and secondary prevention of venous thromboembolic disease, and for prevention of stroke in atrial fibrillation: systematic review, network meta-analysis and cost-effectiveness analysis

Jonathan AC Sterne, Pritesh N Bodalia, Peter A Bryden, Philippa A Davies, Jose A López-López, George N Okoli, Howard HZ Thom, Deborah M Caldwell, Sofia Dias, Diane Eaton, Julian PT Higgins, Will Hollingworth, Chris Salisbury, Jelena Savović, Reecha Sofat, Anya Stephens-Boal, Nicky J Welton and Aroon D Hingorani



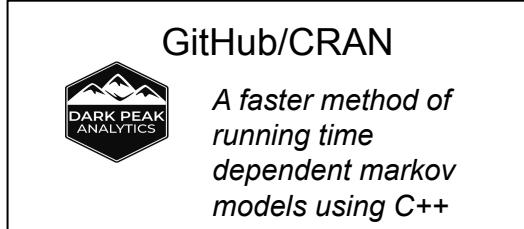
<https://github.com/feralaes/cdx2cea>

<https://github.com/Bogdasayen/DOACs-AF-Economic-model>



Application to HEDS: Software development

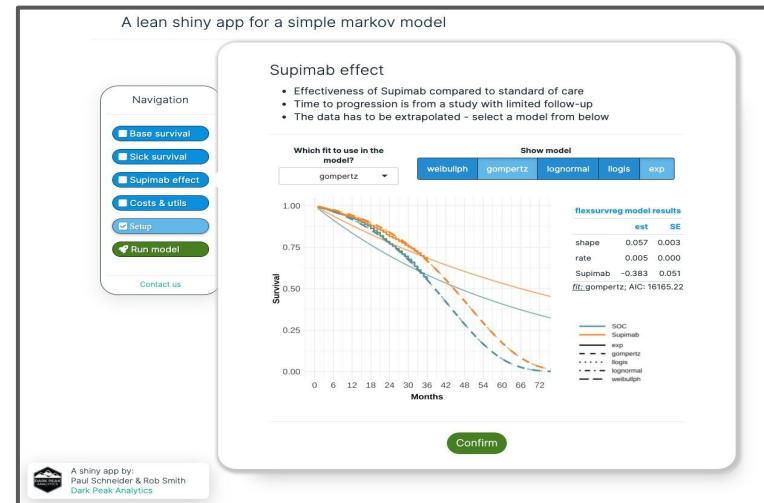
`markov_trace = darkpeak::ArmaTDMarkovLoop(markov_trace, params$trans_mat[, , 3])`



GitHub/CRAN

*A faster method of
running time
dependent markov
models using C++*

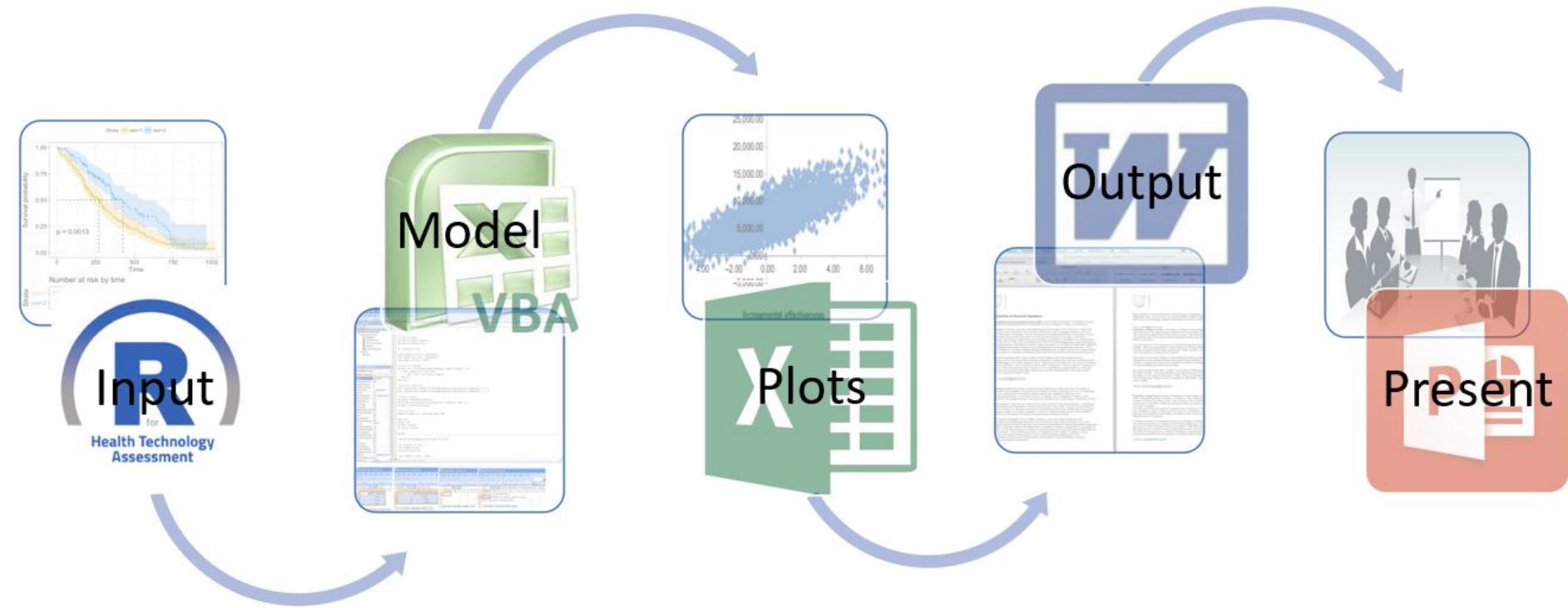
<https://github.com/dark-peak-analytics/darkpeak>



<https://darkpeakanalytics.shinyapps.io/sadm-mk2/>

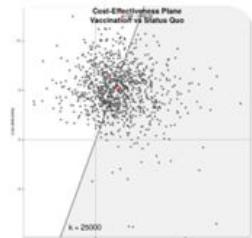
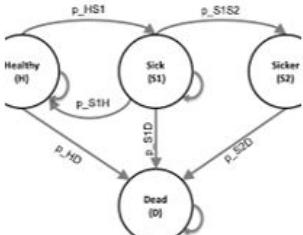
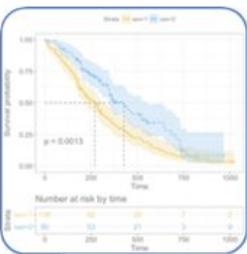


Application to HEDS: Current Process





Application to HEDS: Future Process



VALUING HEALTH TECHNOLOGIES AT NICE:
RECOMMENDATIONS FOR IMPROVED INCORPORATION
OF TREATMENT VALUE IN HTA

DAVID GOLDSTEIN, DAVID LARSON, NATHAN P. PHILIPSON*, AND WOLFGANG SONNENBERG
*Department of Health Services and Radiation, RAND Corporation, Santa Monica, CA, USA
†Department of Health Services, University of Southern California, Los Angeles, CA, USA
‡Department of Health Services, University of Southern California, Los Angeles, CA, USA

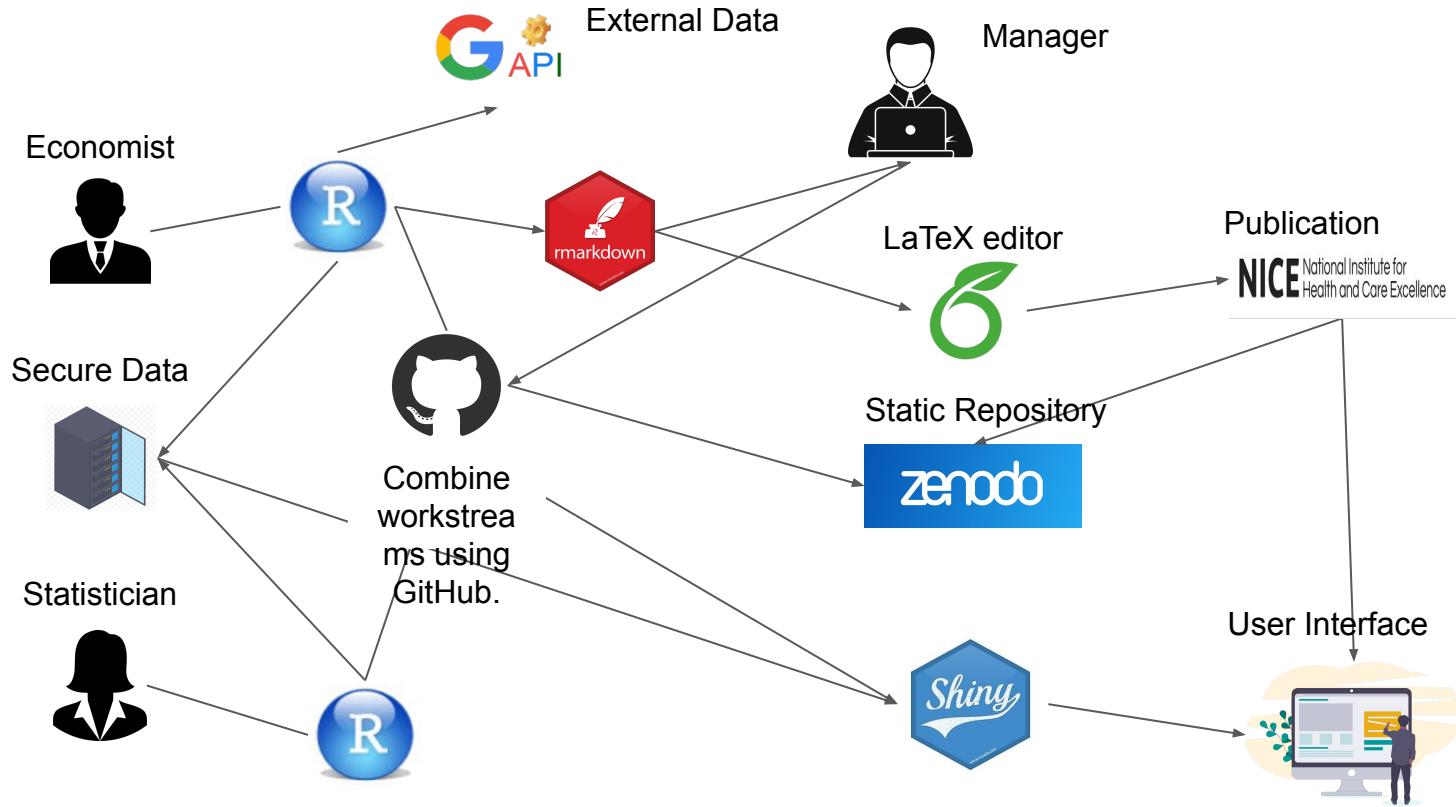
1. INTRODUCTION

Reimbursement and coverage decisions for medical technologies and services are complex negotiations between payers, manufacturers, and providers of medical technologies. In the UK, the National Institute for Health and Care Excellence (NICE) is responsible for assessing the value of new medical technologies, and making recommendations that guide National Health Service (NHS) commissioners of health care to make decisions about the availability of new treatments. One of the key challenges for NICE in its role as a technology appraiser is how to incorporate the concept of 'treatment value' into its decision-making process. The UK has a long history of valuing health care technologies, and could benefit from what could be learned if similar NICE assessments were to be implemented there. This paper presents a study of valuing health care technologies in the UK and provides a strategy for incorporating treatment value into NICE's decision-making process. It is hoped that this paper will assist in a greater understanding of how establishing the value of a technology can be particularly challenging for a committee made up of health economists, health care providers, and health care consumers who are asked to prioritize their choices in a 'cost per QALY' value. As part of this process, we were asked to present our views on what a 'cost per QALY' value and how this could be incorporated in decisions within the NHS.



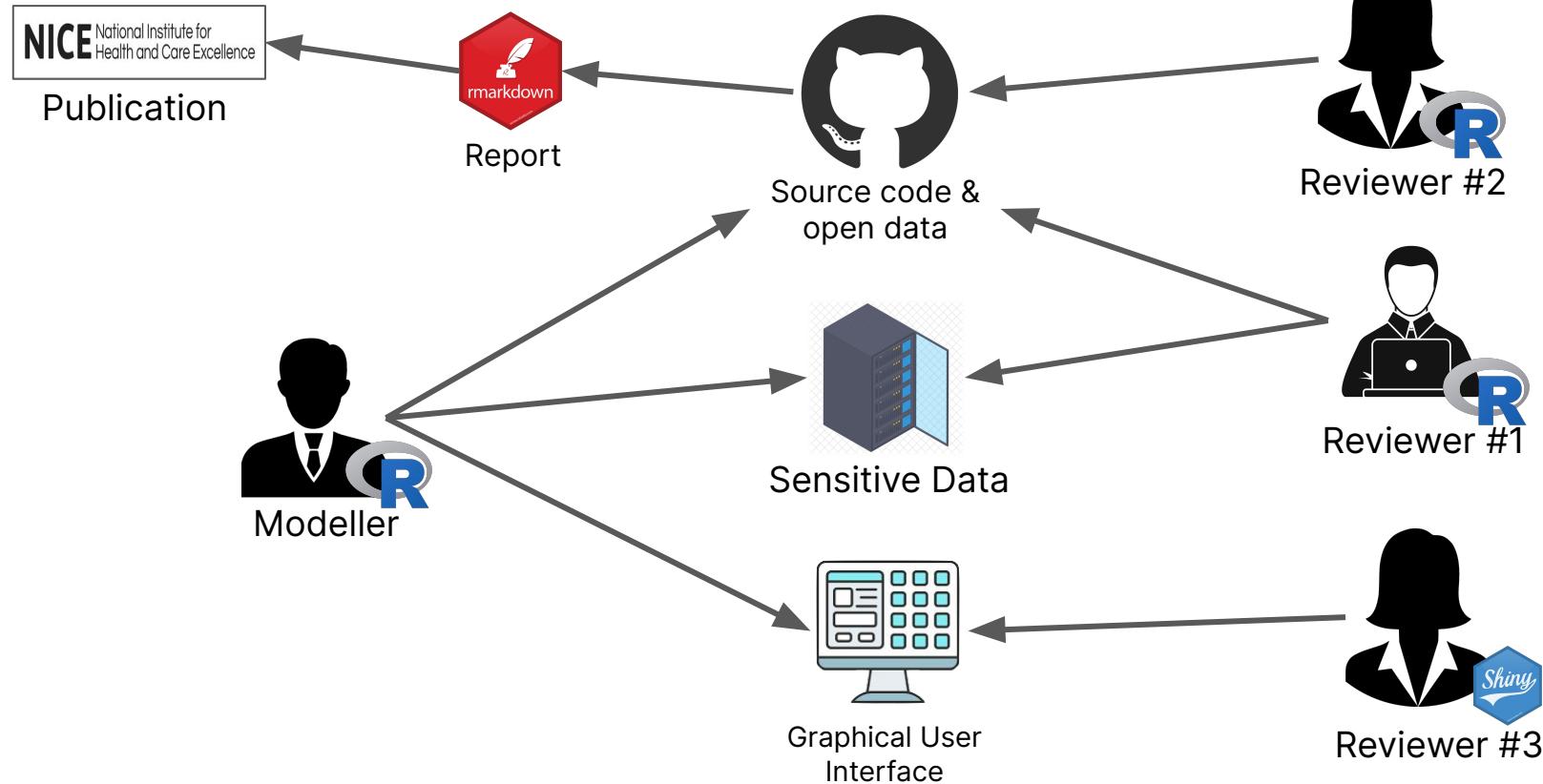


Application to HEDS: Example case study



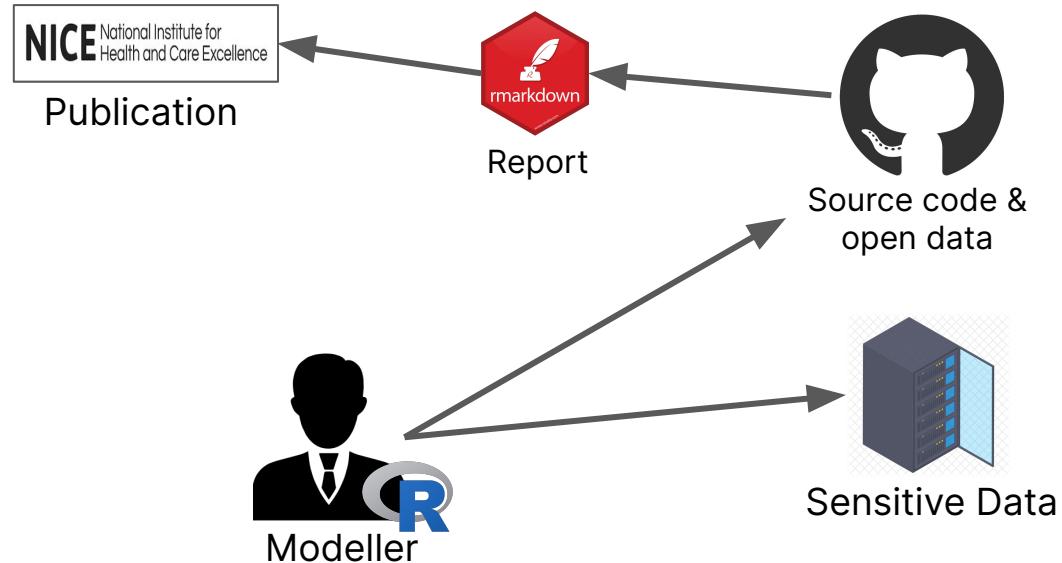


Application to HEDS: External Review options



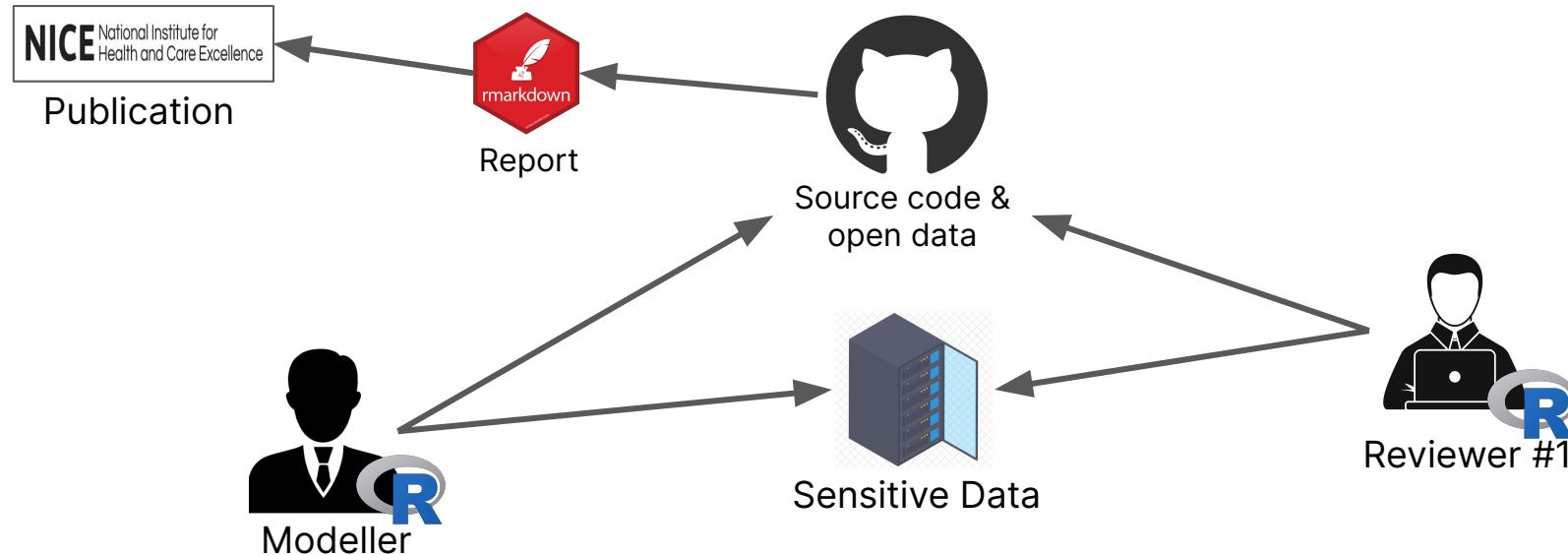


Application to HEDS: External Review options



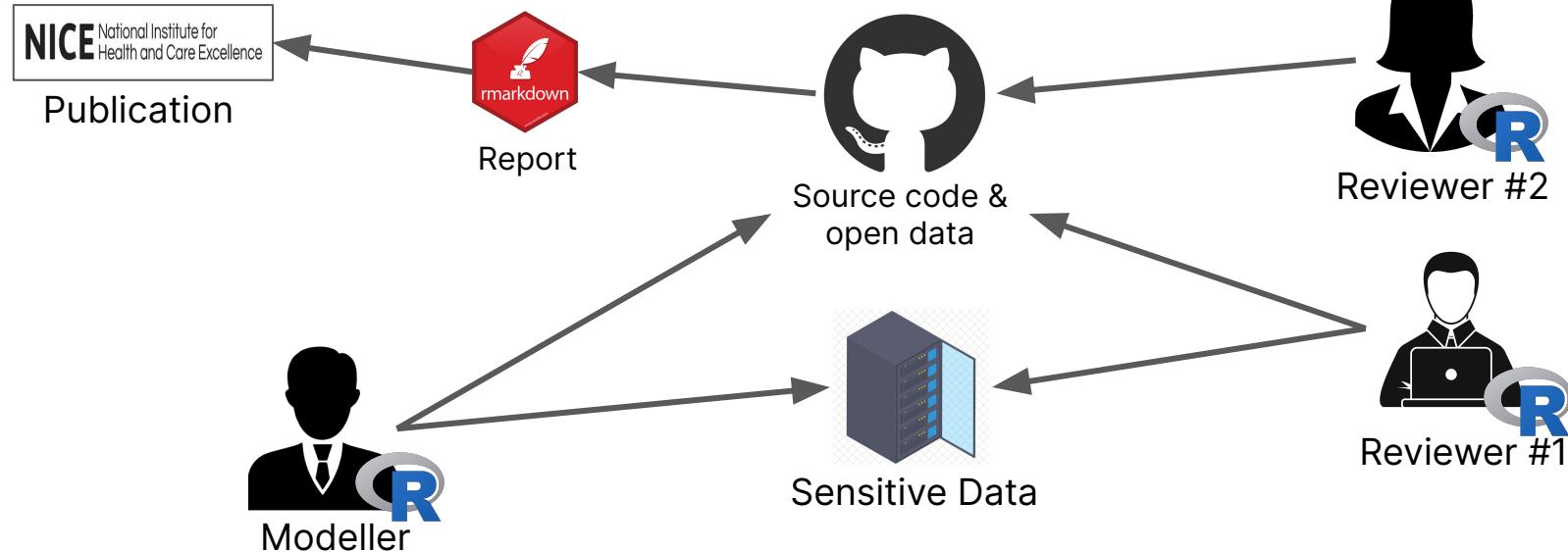


Application to HEDS: External Review options



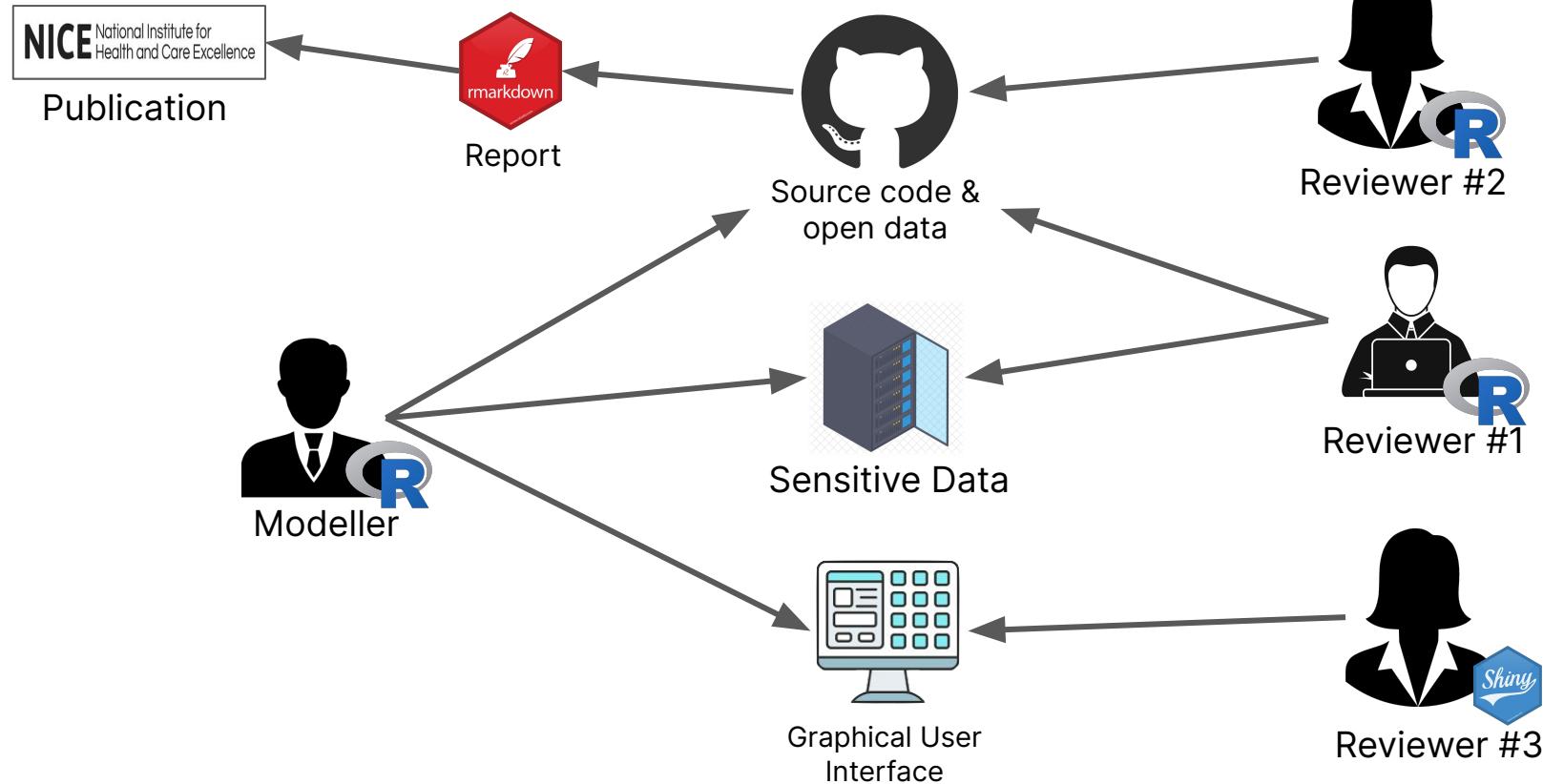


Application to HEDS: External Review options





Application to HEDS: External Review options





Session 2 - Demonstration

*Working together using
RStudio, Git and GitHub*



University of
Sheffield





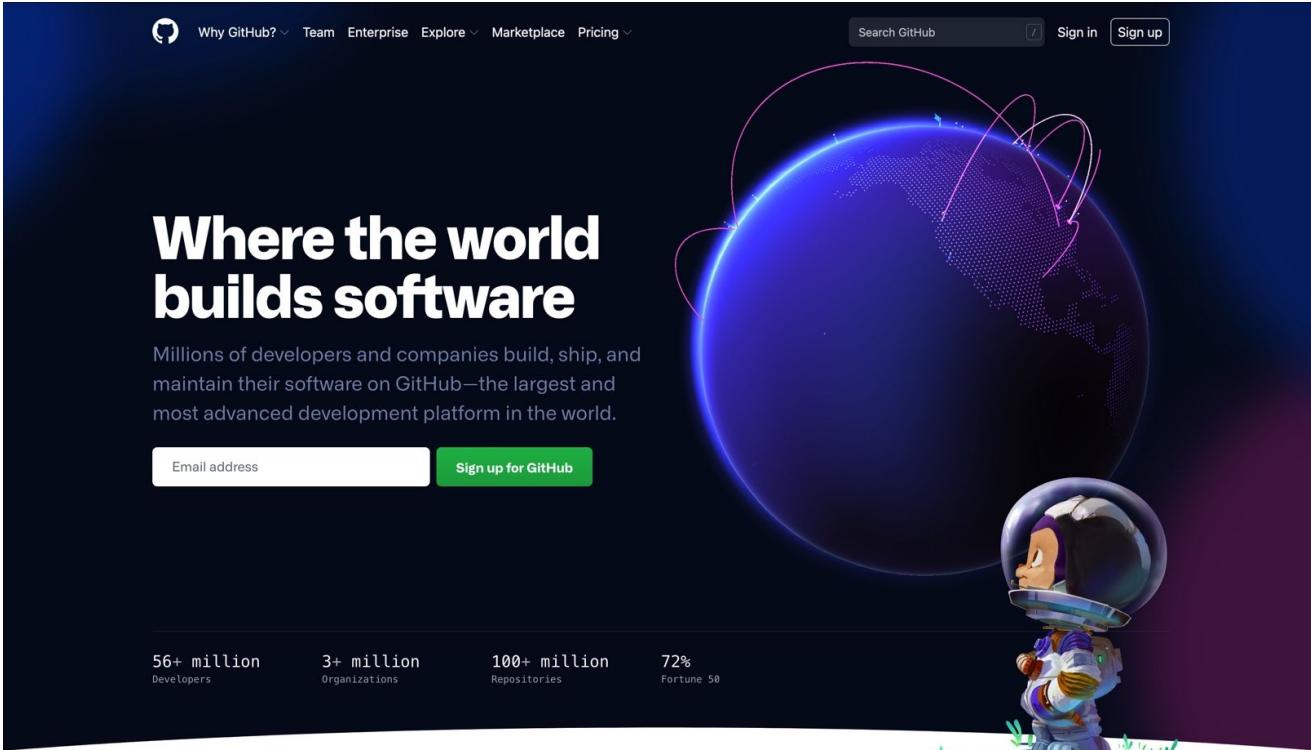
Working on GitHub

What we will cover in this session:

- Repositories - setting up your first repository...
- Branches - creating your first branch, and pull request into ‘main’
- Forks - creating a fork of an existing repository.
- Merging - merging changes made in one branch/fork into another.
- Issues - creating an issue where an error is identified or for a new task
- Discussions - contributing to a discussions.
- Projects - using kanban boards. (Ignore for now).



Signing up to GitHub



The image shows the GitHub homepage. At the top, there is a navigation bar with links for "Why GitHub?", "Team", "Enterprise", "Explore", "Marketplace", and "Pricing". On the right side of the header are "Search GitHub", "Sign in", and "Sign up" buttons. The main visual is a large, glowing blue and purple globe representing Earth, with several pinkish-purple curved lines forming a network or path around it. In the bottom right corner, there is a small, colorful cartoon illustration of an astronaut in a space suit standing on a rocky surface with some green plants growing nearby. The central text on the page reads "Where the world builds software" in large white letters, followed by a smaller paragraph: "Millions of developers and companies build, ship, and maintain their software on GitHub—the largest and most advanced development platform in the world." Below this text are two input fields: a white one labeled "Email address" and a green one labeled "Sign up for GitHub". At the bottom of the page, there are four statistics: "56+ million Developers", "3+ million Organizations", "100+ million Repositories", and "72% Fortune 50".



hello-world: first repository

1. In the upper-right corner of any page, use the drop-down menu, and select '**New repository**'.
2. In the Repository name box, enter '**hello-world**'.
3. In the Description box, write a short description.
4. Select whether your repository will be **Public** or **Private**.
5. Select '**Add a README**' file.
6. Leave *.gitignore template* and *Licence* on **None**
7. Click '**Create repository**'.

<https://docs.github.com/en/get-started/quickstart/hello-world>



Raise your hand when you are finished...

Create a new repository

A repository contains all project files, including the revision history. Already have a project repository elsewhere?
[Import a repository.](#)

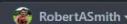
 Single sign-on to see more options within the test-and-trace organization.

Repository template

Start your repository with a template repository's contents.

No template ▾

Owner *



RobertASmith

Repository name *



/

Great repository names are short and memorable. Need inspiration? How about [upgraded-funicular](#)?

Description (optional)

 Public

Anyone on the internet can see this repository. You choose who can commit.

 Private

You choose who can see and commit to this repository.

Initialize this repository with:

Skip this step if you're importing an existing repository.

Add a README file

This is where you can write a long description for your project. [Learn more.](#)

Add .gitignore

Choose which files not to track from a list of templates. [Learn more.](#)

.gitignore template: None ▾

Choose a license

A license tells others what they can and can't do with your code. [Learn more.](#)

License: None ▾



hello-world: your first commit

1. Once the repository is created, **commit a change to the README file** with your name.

<https://docs.github.com/en/get-started/quickstart/hello-world>



Raise your hand when you are finished...



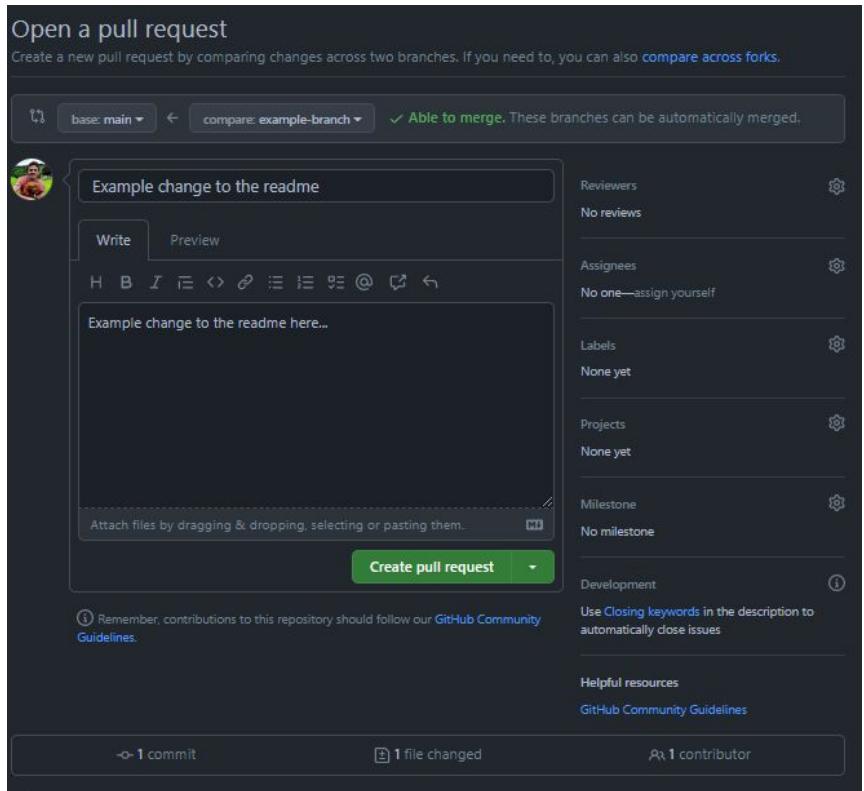
hello-world: branches and pull requests

1. Create a new branch called **example-branch**.
2. Make a minor change to the **README** file - maybe add a preferred name.
3. Create a *pull request* to the **main** branch.
4. Review the *pull request & merge*.

Note: reviewers, assignees, projects, labels etc...



Raise your hand when you are finished...



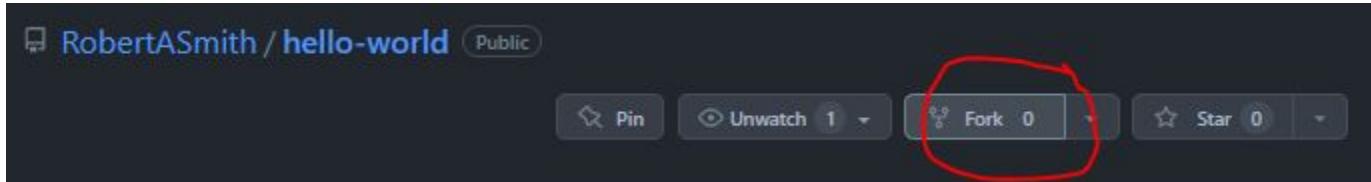
The screenshot shows the GitHub interface for opening a pull request. The title bar says 'Open a pull request' and indicates the base is 'main' and the compare is 'example-branch'. A green checkmark says 'Able to merge. These branches can be automatically merged.' The main area has a 'Write' tab selected, showing a rich text editor with placeholder text 'Example change to the readme here...'. To the right, there are sections for 'Reviewers' (no reviews), 'Assignees' (no one assigned), 'Labels' (none yet), 'Projects' (none yet), 'Milestone' (no milestone), and 'Development' (use closing keywords). At the bottom, there's a note about GitHub Community Guidelines and a 'Create pull request' button.



hello-world: fork and pull request

1. Fork the **RobertASmith/hello-world** repo.
2. Add a file, using 'Add file>Create New File'. The file should be called *<your_initials>_example.R*
3. Create a pull request to *RobertASmith/hello-world* 'main' branch.

Note: Make sure you add a nice comment on the pull request.

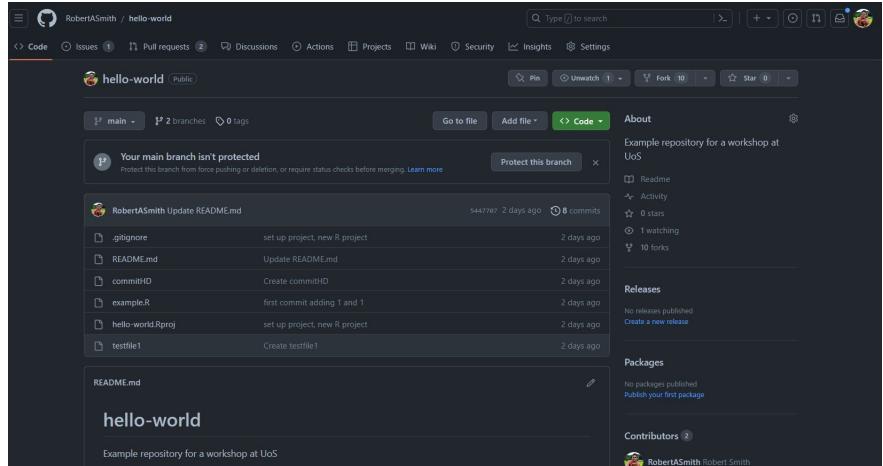


Raise your hand when you are finished...



Issues, discussions and projects

- How to create a new issue
 - Assign - who is this issue for
 - Label - do any apply?
 - Line references - from code or with
```
- Discussions (Q&A)
  - Start a discussion
  - Reference code using ```
  - Like & respond
- Projects
  - Everyone loves a kanban board



The screenshot shows a GitHub repository named "RobertaASmith / hello-world". The repository is public and contains 5 commits by RobertaASmith. The commits are:

- Update README.md (5447787, 2 days ago)
- .gitignore (set up project, new R project, 2 days ago)
- README.md (Update README.md, 2 days ago)
- commitID (Create commitID, 2 days ago)
- example.R (first commit adding 1 and 1, 2 days ago)
- hello-world.Rproj (set up project, new R project, 2 days ago)
- testfile1 (Create testfile1, 2 days ago)

The repository has 10 forks and 0 stars. It is a "Example repository for a workshop at UoS".

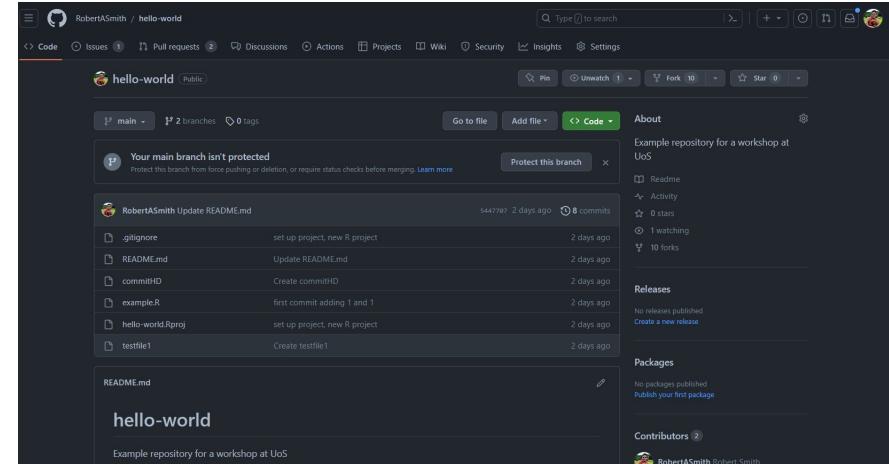


Raise your hand when you are finished...



# Further content (if there is time)

- 1) Create a new issue referencing a line of the readme in the repository.
- 2) Add an R script called ‘<your\_initials>\_example.R’ with a simple calculation in it: ‘ $2+2 == 5$ ’



The screenshot shows a GitHub repository page for 'RobertaSmith / hello-world'. The repository is public and has 2 branches, 0 tags, and 8 commits. The main branch is not protected. The README.md file has been updated. The repository is described as an 'Example repository for a workshop at UoS'.

Repository details:

- Owner: RobertaSmith
- Name: hello-world
- Status: Public
- Branches: 2
- Tags: 0
- Commits: 8
- Last commit: 2 days ago
- Protected branch: No

Commit history:

- Update README.md (by RobertaSmith, 2 days ago)
- .gitignore (by gitignore, 2 days ago)
- README.md (by RobertaSmith, 2 days ago)
- commitID (by commitID, 2 days ago)
- example.R (by example.R, 2 days ago)
- hello-world.Rproj (by hello-world.Rproj, 2 days ago)
- testfile1 (by testfile1, 2 days ago)

README.md content:

```
hello-world
```

Description: Example repository for a workshop at UoS

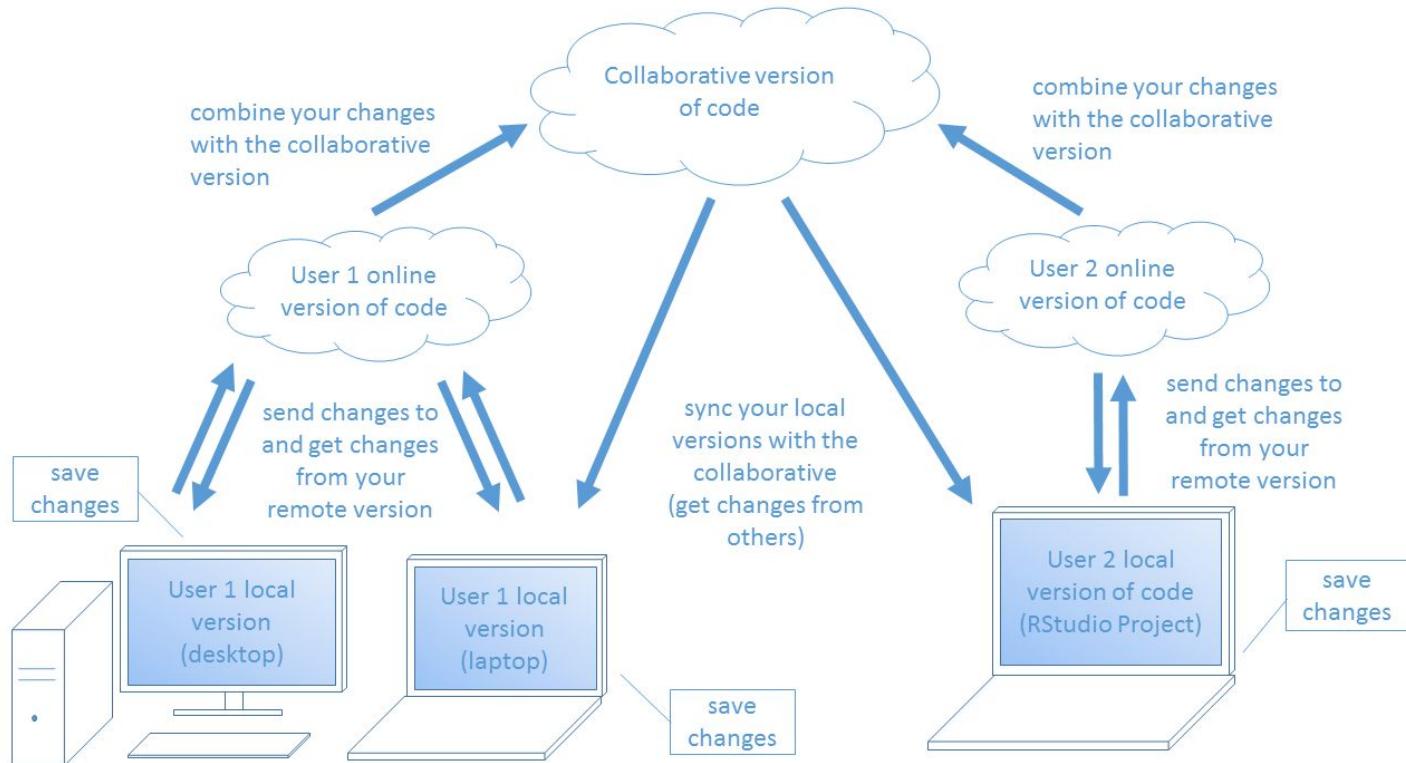


# Setting up Git and RStudio

*Working alone  
from RStudio*



# Setting up GitHub and RStudio





Before we go into details:

- Git and GitHub are difficult.
- It's a different type of skill to statistics & health economics.
- It's very hard to learn by reading/listening to me, but this course is the perfect time to learn.
- Don't worry if you don't understand at first.
- It will become clear.
- Keep making lots of small mistakes.
- **IT IS REALLY WORTH IT!**





# Setting up Git and RStudio

Now - the tricky bit - we are going to move through this slowly ... following the coursebook section:

<https://r4he-pentag.netlify.app/version#getting-a-github-account>

- 1) **Confirm Git is installed**, type `git --version` into your command prompt (or mac equivalent).

If Git is not already installed then [install it](#), alternative guide [here](#).

- 2) **Link Git to RStudio** in `Tools > Global Options > Git/SVN`

Git location can be found by typing `where git.exe` into the command prompt. If struggling see alternative guide [here](#)

- 3) **Set up a GitHub Personal Access Token (PAT)**

This can be done on [GitHub](#). If struggling see alternative guide [here](#)



Raise your hand when you are finished...

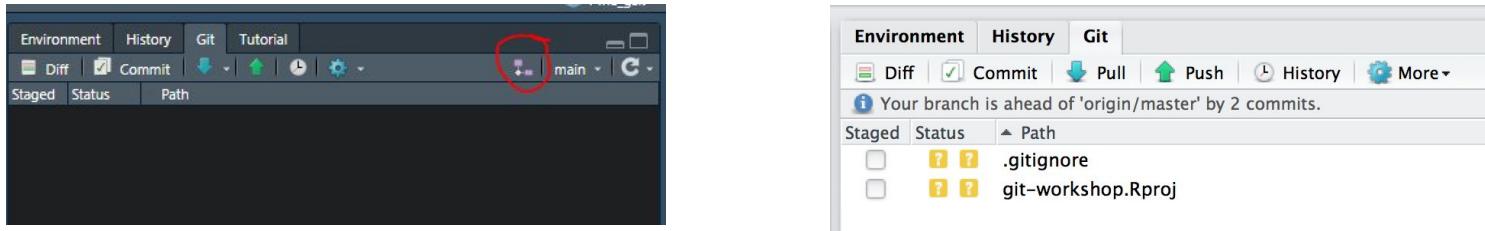


# Cloning a GitHub Repo from RStudio

4) Clone the repository we created earlier using RStudio by creating a new project from a version controlled directory - this should force you to enter your *PAT* we created in the last step.

<https://r4he-pentag.netlify.app/version#cloning-the-repository-with-rstudio>

5) Create a new branch in RStudio, and then 1) add a file, 2) commit locally and 3) push to GitHub.



6) Go to GitHub and create a pull request from your branch into the ‘main’ branch.



Raise your hand when you are finished...

# Giting together

Since there are lots of us this a good chance to practice ...

- 1) In RStudio - go to your clone or fork of the repository I created earlier.
- 2) Go to 'main' and *pull* to ensure you are up to date.
- 3) Create a new *branch* called *<initials>\_branch*
- 4) Create a new file called *<initials>\_example1.R*
- 5) Write something interesting in line 1. Saving the file once you have finished.
- 6) When you have saved the file - *commit* and *push* to your fork.
- 7) Go to GitHub and send a *pull request* from your fork to the main repo.
- 8) Have a look at what others are doing, can you see changes to the repo live.



Raise your hand when you are finished...

## Branches and Reviewers



# Question & Answer

# ChatGPT, write me a haiku about Git ...

Code is my art,  
Crafting lines with care and heart,  
Git keeps it safe.



University of  
Sheffield

Short course created by Dr. Robert Smith, University of Sheffield & Dark Peak Analytics, for: OPENFEST – 2023



# Thank you

Talks: <https://github.com/RobertASmith/talks>

Journal Publications: <https://scholar.google.com/citations?user=nEJ0x6IAAAAJ&hl=en>

Source Code: <https://github.com/RobertASmith>; <https://github.com/orgs/dark-peak-analytics>

Dark Peak Analytics Website: <https://darkpeakanalytics.com/>

---

Dr. Robert Smith

Director, Dark Peak Analytics, [rsmith@darkpeakanalytics.com](mailto:rsmith@darkpeakanalytics.com)

Wellcome Funded Post-Doc, ScHARR, [robert.smith@sheffield.ac.uk](mailto:robert.smith@sheffield.ac.uk)

